

# Anharmonic Torsional Stiffness of DNA Revealed under Small External Torques

Alexey K. Mazur\*

CNRS UPR9080, Institut de Biologie Physico-Chimique,  
13, rue Pierre et Marie Curie, Paris, 75005, France.

DNA supercoiling plays an important role in a variety of cellular processes. The torsional stress related with supercoiling may be also involved in gene regulation through the local structure and dynamics of the double helix. To check this possibility steady torsional stress was applied to DNA in the course of all-atom molecular dynamics simulations. It is found that small static untwisting significantly reduces the torsional persistence length ( $l_t$ ) of GC-alternating DNA. For the AT-alternating sequence a smaller effect of the opposite sign is observed. As a result, the measured  $l_t$  values are similar under zero stress, but diverge with untwisting. The effect is traced to sequence-specific asymmetry of local torsional fluctuations, and it should be small in long random DNA due to compensation. In contrast, the stiffness of special short sequences can vary significantly, which gives a simple possibility of gene regulation via probabilities of strong fluctuations. These results have important implications for the role of local DNA twisting in complexes with transcription factors.

PACS numbers: 87.14.gk 87.15.H- 87.15.ap 87.15.ak

The double helical DNA in living cells is subjected to a constitutive unwinding torque created by special enzymes. This forces DNA to fold in a supercoiled state similarly to a flexible rod with bending and twisting elasticity. The supercoiling is long known to play an important role in a variety of cellular processes [1]. Its magnitude changes regularly during the cell cycle and in response to environmental conditions, which is accompanied by activation or suppression of certain genes [2]. In *E. coli*, relaxation of the superhelical stress simultaneously alters activity of 306 genes (7% of the genome), with 106 genes activated and other deactivated [3]. The genes concerned are functionally diverse, widely dispersed throughout the chromosome, and the effect is dose-dependent. These and many similar observations suggest that the DNA supercoiling is used as a universal transcriptional regulator [2], but the corresponding physical mechanisms are not clear.

Detailed studies indicate that the promoter sensitivity to supercoiling stems from the recognition of promoter elements by RNA polymerase, and that it does not require DNA melting or transitions to alternative forms [4]. The supercoiling torque is distributed between twisting and writhing so that the untwisting of the double helix is estimated as 1-2% [5], which is below the thermal noise and too small for reliable recognition. However, the action of the torsional stress can be conveyed through a property rather than the structure of the double helix. The behavior of the supercoiled DNA is governed by the interplay between the local bending and twisting fluctuations. If the bending flexibility or the torsional stiffness of the double helix vary with forced untwisting, parameters of thermal fluctuations could be noticeably affected already for short DNA stretches involved in the recognition. This idea is appealing and it is supported by some earlier data for long DNA [6-8]. Local torsional fluctuations are likely to be involved in regulation directly. In

bacterial promoters, the optimal linker between the -10 and -35 elements involves 16 base pair steps (bps), but in promoters sensitive to supercoiling it is usually one step shorter or longer [4, 9]. One step corresponds to rotation by  $34.5^\circ$ , which approximately equals the root-mean square width of torsional fluctuations for the linker. Very strong torsional fluctuations of short DNA stretches are necessary for activation of some animal promoters [10].

Local effects of the torsional stress are difficult to reveal experimentally, but they can be probed by all-atom MD simulations. New methods were recently developed to apply steady forces and torques to short stretches of DNA [11]. In contrast to twisting by periodic boundary constraints and potential restraints used earlier [12-14] the steady stress emulates local conditions of a short fragment in a long supercoiled DNA, which makes possible evaluation of elastic parameters under very low torsional load corresponding to physiological conditions. This method captures linear elastic responses as well as the twist-stretch coupling effect under small torques corresponding to physiological degree of supercoiling [11]. Here we present the results of the first computational study of the elastic parameters of DNA in such conditions.

Dynamics of two tetradecamer DNA with AT- and GC-alternating sequences, respectively, were simulated in explicit aqueous solution using earlier described protocols [11]. For each duplex, nine 164 ns trajectories of all-atom dynamics were computed with fixed torque values in the range  $\pm 20$  pN·nm, which gives about  $3 \mu\text{s}$  of simulations in total. Three additional trajectories were computed for the GC-alternating fragment for verification. Below we consider only evaluation of the torsional stiffness. Other methods and protocols are described in Appendix. In the harmonic approximation the torsional free energy of a DNA fragment of length  $L$  subjected to external torque

$\tau$  is

$$U(\Phi) = kT \frac{l_t}{2L} (\Phi - \Phi_\tau)^2 \quad (1)$$

where  $\Phi$  is the overall winding angle,  $\Phi_\tau$  is its equilibrium value,  $l_t$  is the torsional persistence length, and  $kT$  is the Boltzmann's factor. The equilibrium winding varies with the torque as

$$\Phi_\tau - \Phi_0 = \frac{\tau L}{kT l_t} \quad (2)$$

In the course of MD simulations one measures the probability distribution  $P_\Phi$  for the winding angle of one helical turn which, in the limit of infinite sampling, has a canonical form

$$P_\Phi \sim \exp \left[ -\frac{l_t}{2L} (\Phi - \Phi_\tau)^2 \right]. \quad (3)$$

The equilibrium winding is estimated as the time average  $\langle \Phi \rangle_t$ , and the torsional persistence length  $l_t$  is extracted from the time variance  $\Delta_t^2 \Phi$ . The potential of mean force (PMF) corresponding to any Gaussian distribution is quadratic, but if the harmonic approximation is truly valid,  $l_t$  must be constant with different  $\tau$ .

The top panel of Fig. 1 shows variations of  $\Phi_\tau$  corresponding to Eq. (2). All measurements were taken for the central 12 bp stretches, with the two terminal steps ignored, which gives about one helical turn. The amplitude of the forced winding is  $\pm 2\%$ , i.e. about  $0.7^\circ$  per base pair. The straight lines shown have the slopes corresponding to  $l_t$  obtained under zero torque. In the range of torques  $\pm 10$  pN·nm the points are compatible with a linear elastic response (harmonic elasticity). Beyond this range the profile remains roughly linear for the AT-alternating sequence, but for the GC-alternating duplex evident deviations from harmonicity are found. These deviations are reproducible and quite strong. If the  $l_t$  value were evaluated by Eq. (2) using  $\Phi_\tau$  for  $\tau = \pm 20$  pN·nm it would be about 200 nm.

The measured torsion persistence length changes with the applied torque as shown in the bottom panel. The GC-alternating sequence exhibits strong anharmonicity, with the twist increase of  $1.4^\circ$  per bps accompanied by 30% growth in  $l_t$ . For the AT-alternating sequence, the  $l_t$  profile is nearly flat with a small decreasing trend. This trend becomes more visible with stronger twisting (article in preparation). The bending stiffness varies somewhat beyond the estimated statistical errors, but without regular trends.

Fig. 2 shows the probability distributions  $P_\Phi$  for the GC-alternating sequence for three representative values of  $\tau$ . All of the distributions are close to the analytical Gaussians defined by Eq. (3) with different  $l_t$ . Since the width of the bells changes, the neat shapes of the computed distributions are not due to the harmonicity

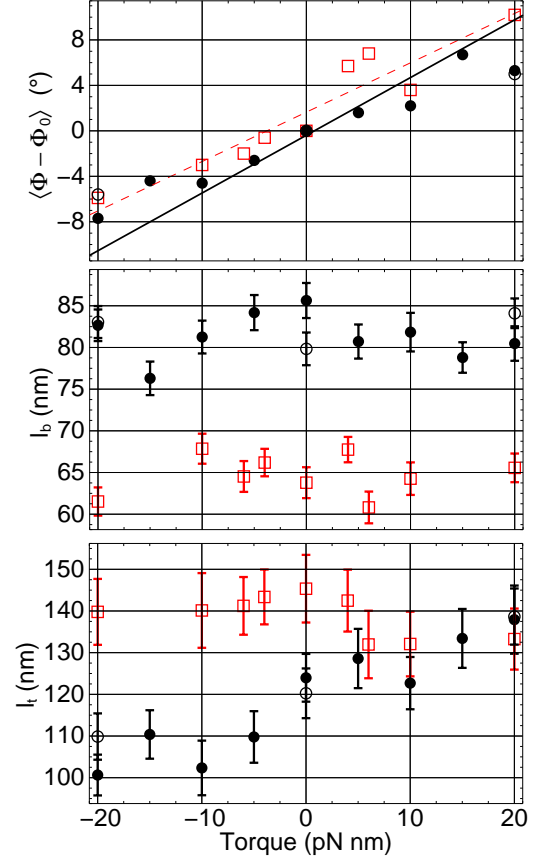


FIG. 1 Color online. Representative torque dependences obtained by all-atom MD simulations. The results are shown for the overall twisting (top panel), the bending persistence length ( $l_b$ , middle panel), and the torsional persistence length (bottom panel) of the AT-alternating ( $\Phi_0 \approx 363.1^\circ$ , red squares) and GC-alternating ( $\Phi_0 \approx 381.8^\circ$ , black circles) sequences. The open circles feature the verification tests. The straight lines on the top panel correspond to Eq. (2) with  $l_t = 124$  nm (solid black line) and  $l_t = 145$  nm (dashed red line). The error bars show statistical errors evaluated by the method of block averages (see Appendix). In the top panel the symbol size corresponds to maximal errors.

of the torsional potential. These Gaussian shapes result from the central limit theorem of the probability theory whatever the underlying potential is. As seen in Fig. 3, the single-step twist fluctuations at GpC and CpG steps produce wide and skewed non-Gaussian distributions strongly different from that predicted by Eq. (1) (see Appendix). With the temperature around 300K, the local DNA dynamics goes far beyond the area where the harmonic approximation is valid. However, the torsional fluctuations of four consecutive bps already give an almost ideal Gaussian. It can be formally described by Eq. (1) and (3), but the shape of this bell does not correspond to the harmonic approximation of the local free energy. The Gaussian profile of fluctuations in long DNA is linked with the single-step distributions by a linear growth of the variance with the chain length. Consequently, not just the apex zones of the skewed distributions in Fig. 3,

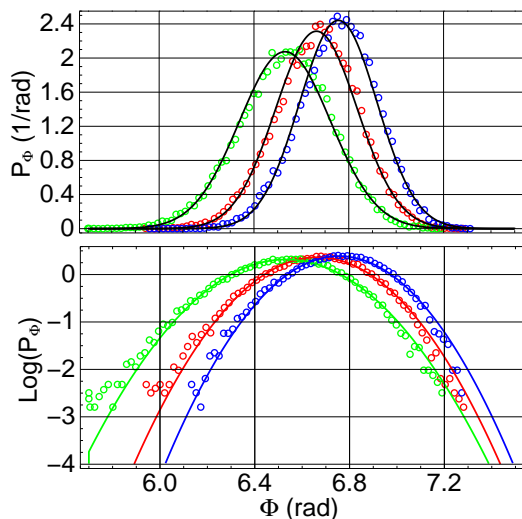


FIG. 2 Color online. The normalized probability density  $P_\Phi$  obtained with different applied torques. From left to right, the MD results are shown for  $\tau = -20, 0$ , and  $+20$  pN·nm by green, red, and blue points, respectively. The solid lines exhibit analytical distributions Eq. (3) corresponding to the measured values of  $l_t$  and  $\Phi_\tau$ . The lower panel displays the same data in semi-logarithmic coordinates.

but their entire shapes contribute. Therefore, the anharmonicity is significant, but hidden. In addition, the twist fluctuations at consecutive steps are anticorrelated and partially cancel out.

The asymmetry of the single-step PMFs is the probable cause of the variable torsional stiffness of the GC-alternating fragment. In the first approximation, the  $l_t$  value is proportional to the second derivative of the PMF in the energy minimum (see Eq. (1)). For an asymmetric PMF a decrease in  $l_t$  may be expected when the external torque pushes towards the even slope of the energy profile. In the GC-alternating sequence both single step distributions are left-skewed (see Fig. 3); so the right-hand slope of the PMF is steeper than the opposite one, which explains the sign of the trend in  $l_t$  observed in Fig. 1. The nearly flat  $l_t$  profile for the AT-alternating fragment can be also rationalized because in this case a strong positive skewness of TpA steps is partially compensated by a negative skewness of ApT steps (see Appendix). Preliminary analysis of other sequences reveals that the strong negative skewness of the CpG single-step distributions is exceptional (see Appendix). The homopolymer ApA and GpG steps are nearly symmetrical whereas the single-step distributions for AG- and AC-alternating DNA indicate that they would behave similarly to the AT-alternating fragment. These conclusions should be yet verified in more intensive computations, but we expect that for random DNA the macroscopic torsional stiffness should be nearly constant because among the steps with skewed distributions positive and negative skewness are equally represented. In contrast, for short sequence motives an-

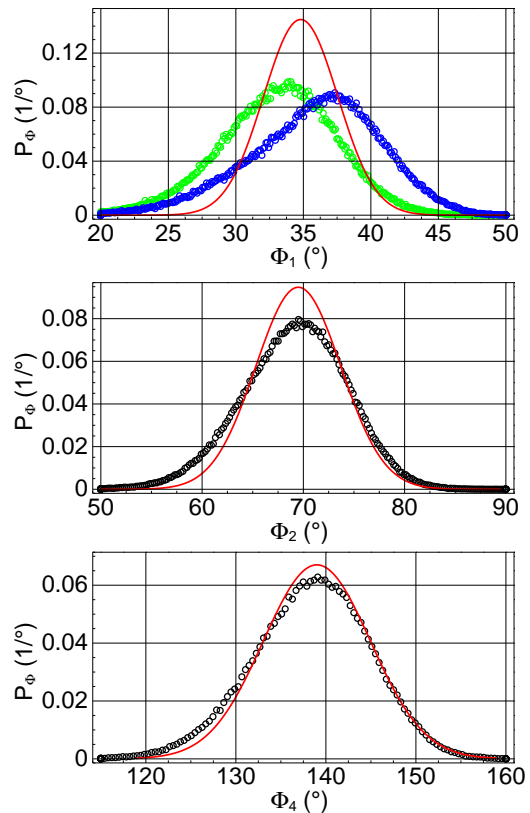


FIG. 3 Color online. The probability density  $P_\Phi$  for GC-alternating fragments of one, two and four bps (from top to bottom) obtained with  $\tau = 0$ . The solid red lines exhibit the analytical distributions Eq. (3) corresponding to the measured values of  $l_t$  and  $\Phi_0$ . On the top panel, the distributions for GpC and CpG steps are shown in green (left) and blue (right), respectively.

harmonic effects of both signs are possible. They can be very significant because biological systems operate with much larger torques than we use here. For instance, the binding sites of the phage 434 repressor contain a variable 4 bp spacer that does not interact with the protein and supposedly participates in gene control via the sequence-dependent elasticity [15]. In the complexed state, this spacer is always overtwisted by about  $30^\circ$  [16], that is ten times the amplitude of twisting in Fig. 1.

The experimental bending rigidity of free DNA is characterized by  $l_b \approx 50$  nm [17]. The measured  $l_t$  values vary between 36 and 109 nm depending upon the specific methods and conditions [18]. Observations of sequence effects are rare [19], and there are a few reports on the influence of supercoiling [6–8]. If we assume that MD overestimates the stiffness of DNA uniformly then the convergent estimate of  $l_t$  is around 90 nm, close to its value in single molecule experiments [20, 21]. The bias can be due to the neutralizing salt condition in MD or other factors (see Appendix). The nearly quantitative agreement between MD and experiment is remarkable because none of the parameters used in simulations was adjusted to re-

produce the DNA elasticity. We hope, therefore, that the detailed microscopic picture provided by MD captures the qualitative physical trends dictated by the atom-level mechanics of the double helix. Our results argue that, under normal temperature, the local DNA elasticity is strongly anharmonic. Extrapolation from the apparent harmonic behavior of macroscopic DNA is not justified despite a good agreement with atomistic simulations for chain lengths beyond one helical turn [22, 23]. In addition, these computational observations shed new light upon some earlier controversial issues.

According to Fig. 1, with the helical twist slightly shifted from the equilibrium value the sequence dependence of the DNA elasticity can be significantly changed and enhanced. The measured torsional stiffnesses are similar without applied torque, but diverge with untwisting. The deformability of DNA is long considered as a possible governing factor in the sequence-specific site recognition [15], but this mechanism requires strong sequence dependence of elastic parameters compared to that observed in experiments with free DNA [19]. As we see the properties of the relaxed DNA cannot be simply transferred to supercoiled and/or protein bound DNA states. Additional studies are necessary to check if the elastic properties of the specific binding sites change under torsional stress. Its magnitude may be very large in some protein complexes [16].

Another debated issue concerns the mechanisms of gene regulation via DNA supercoiling [2, 3]. Many of such observations are readily rationalized if we assume that the sensitive promoters are regulated via the torsional stiffness. Even a slight shift in its value has a dramatic effect on the probabilities of strong twisting fluctuations. Many transcription factors are designed to bind the double helix at two sites separated by a spacer of several base pair steps. They can work as sensors of torsional fluctuations in DNA. A strong twisting fluctuation may be necessary for binding such factor or for recognition by other proteins of a permanently bound torsional sensor. Fig. 2 shows that for fluctuations observable during 164 ns, physiological modulations of the torsional stress would change the corresponding probabilities by several times. For less frequent larger fluctuations the effect would be much stronger. One can extrapolate the pattern in Fig. 2 to events observable in the millisecond time range, and this leads to essentially all-or-nothing switching.

The external torque shifts the distributions in Fig. 2 by changing symmetrically the energies of opposite fluctuations. If the shape of the distributions does not change each pair of curves should give a single intercept between the corresponding two apexes. However, if the shifting is accompanied by widening, one more intercept should appear in the range of large twisting opposite to the torque direction. For instance, the negative torque shifts the distribution in Fig. 2 to the left, but the simultaneous widening raises its right wing and, with very large over-

twisting, the left curve should go above the other two. It is seen in Fig. 2 that the vertical difference between the three plots indeed exhibits a reducing trend with large  $\Phi$ . This effect is somewhat paradoxical and it qualitatively contradicts the behavior of simple models where the torsional energy depends upon a single variable. Our attempts to reproduce it in discrete wormlike chains with anharmonic torsion potentials were unsuccessful. However, such behavior is possible, in principle, due to coupling between different degrees of freedom and it requires further studies.

To conclude, it appears that small external torques can significantly alter the torsional stiffness of the double helical DNA. The effect is sequence-dependent, and, under variable degree of supercoiling, different stretches of the double helix can become locally softer or stiffer. This can represent a versatile mechanism of gene regulation via the probabilities of strong twisting fluctuations.

## APPENDIX

### Simulation protocols

Tetradecamer DNA fragments were modeled with AT-alternating (d(AT)<sub>7</sub>) and GC-alternating (d(GC)<sub>7</sub>) sequences. The starting states for classical MD simulations were prepared as follows. The solute in the canonical B-DNA conformation [24] was immersed in a 6.2-nm cubic cell with a high water density of 1.04. The box was neutralized by placing Na<sup>+</sup> ions at random water positions at least 5 Å from the solute. The system was energy minimized and dynamics were initiated with the Maxwell distribution of generalized momenta at low temperature. The system was next slowly heated to 293 K and equilibrated during 1.0 ns. After that the water density was adjusted to 0.997 by removing the necessary number of water molecules selected randomly at least 5 Å from DNA and ions, and the simulations were continued with NVT ensemble conditions. The temperature was maintained by the Berendsen algorithm [25] applied separately to DNA, water, and ions, with a relaxation time of 10 ps. Simulations with external torques started from equilibrated states after a few nanoseconds of free dynamics.

The AMBER98 forcefield parameters [26, 27] were used with the rigid TIP3P water model [28]. The electrostatic interactions were treated by the SPME method [29], with the common values of Ewald parameters, that is 9 Å truncation for the real space sum and  $\beta \approx 0.35$ . To increase the time step, MD simulations were carried out by the internal coordinate method (ICMD), [30, 31] with the internal DNA mobility limited to essential degrees of freedom and rotation of water molecules and internal DNA groups including only hydrogen atoms slowed down by weighting of the corresponding inertia tensors. [32, 33] The

double-helical DNA was modeled with all backbone torsions, free bond angles in the sugar rings, and rigid bases and phosphate groups. The effect of these constraints is insignificant, as was previously checked through comparisons with standard Cartesian dynamics [22, 32]. The time step was 0.01 ps and the DNA structures were saved every 5 ps. All trajectories were continued to accumulate  $2^{15}$  points, i.e. to about 164 ns. To verify the results for d(GC)<sub>7</sub> three additional trajectories were computed with torques  $\tau = -20, 0$ , and  $+20$  pN·nm, respectively. These computations were carried out in parallel on 32 processors starting from independent equilibrated states. In all simulations the B-DNA conformations were well conserved without visible slow trends or accumulated deformations. Fig. 1S shows some standard time plots for two representative trajectories. It is seen that the overall properties of the double helices remain stable and that the DNA structures were well-equilibrated before the beginning of the production runs.

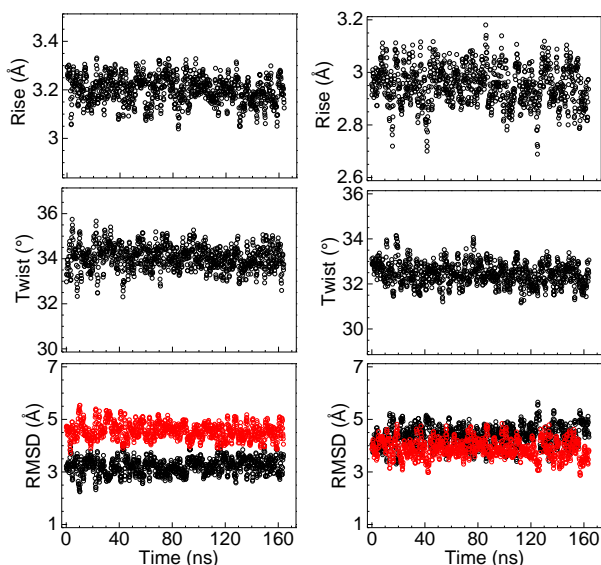


FIG. 1 FIG. 1S: The time evolution of the double helical structures during MD simulations with applied torque  $\tau = -20$  pN·nm. The results for the GC- and AT-alternating sequences are shown in the left and right columns, respectively. The helical parameters are averages for the central 12 bp obtained by program 3DNA [34]. The all-atom RMSDs from the canonical A-DNA (red) and B-DNA (black) were computed for the entire fragment length (14 bp). The analysis was carried out for 1000 states equally spaced throughout the trajectories and all plots were smoothed with a sliding window of 82 ps.

The choice of the fragment length and sequences is consistent with the recent computations [11, 35, 36] and it was dictated by the following considerations. The length slightly larger than one helical turn is convenient for measuring the elastic parameters of DNA [36]. These molecules are homopolymers of AT- and GC-units, therefore, they cannot have distinguished asymmetric structures like static bends. True homopolymer DNA duplexes

have special features and, in free MD with the AMBER forcefield, these structures deviate from the canonical B-DNA stronger than AT- and GC-alternating sequences [37]. The terminal base pairs open rather frequently during nanosecond time scale MD, which significantly perturbs the flanking DNA structure. Because this dynamics cannot be averaged during the accessible duration of MD trajectories, we blocked it by applying non-perturbing upper distance restraints as explained elsewhere [11]. The statistical convergence of MD sampling also suffers from rare transitions of backbone torsion angles to non-standard states considered as forcefield artifacts [38, 39]. In the present simulations such transitions occurred mainly in the terminal base pair steps excluded from analysis. No  $\alpha/\gamma$  flips were observed in the middle d(CG)<sub>6</sub> stretches. A few such transitions that occurred in the d(TA)<sub>6</sub> fragments were documented and discussed in our previous report [11]. They increased the statistical noise in the measured parameters, but did not cause overall structural perturbations. Fig. 2S shows several representative distributions of the backbone torsions involved in non-canonical transitions [38]. The number of  $\alpha/\gamma$  flips in d(TA)<sub>6</sub> was maximum two per trajectory, but some of them were reversed. The distribution with the maximal population of non-canonical states shown in Fig. 2S corresponds to a trajectory with a single stable flip.

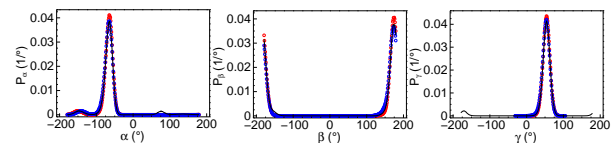


FIG. 2 FIG. 2S: Distributions of  $\alpha$ ,  $\beta$ , and  $\gamma$  backbone torsions for three representative trajectories. Plots shown by open circles correspond to the data in Fig. 1S, with the results for d(TA)<sub>6</sub> and d(CG)<sub>6</sub> in red and blue, respectively. The solid black line shows the distributions for a trajectory with the largest population of non-standard conformers ( $\tau = +4$  pN·nm, d(TA)<sub>6</sub>).

The sampled conformations of the double helix were analyzed by program 3DNA [34], with only 11 central base pair steps considered (d(TA)<sub>6</sub> and d(CG)<sub>6</sub>). Accurate direct measurement of the length of short double helices encounters technical difficulties [36, 40]. Therefore, the corresponding DNA length was computed as  $11 \times 0.335$  nm, that is assuming the length of one step corresponding to experiment. The error in the attributed length can be partially responsible for the systematic bias in the absolute values of other measured parameters, but it does not affect qualitative trends.

### External torques

Steady external torsional load was applied as described in detail in our recent report [11]. This method dis-

tributes forces over selected groups of atoms and compensates them by reactions applied to other atoms so as to zero the total external force and torque. Because the forces are applied at different points internal stress and deformations are introduced that correspond to overall twisting. The method was thoroughly verified in Brownian dynamics simulations of calibrated discrete wormlike chain models [11]. Notably, it was checked that the torque values in the range of interest cause linear elastic responses in perfect agreement with theoretical predictions and negligible side effects.

### Evaluation of statistical errors

Evaluation of errors in MD simulations is based upon the following assertions from the probability theory. Consider a random variable  $x$  with expectation  $\mathbf{M}x = \xi$  and variance  $\mathbf{D}x = \sigma^2$ . We can take  $n$  samples of  $x$  and compute

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{k=1}^n x_k$$

and

$$S^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$$

called the sample average and variance, respectively. Both  $\bar{x}$  and  $S^2$  are random variables, with  $\mathbf{M}\bar{x} = \xi$  and  $\mathbf{M}S^2 = \sigma^2$ , i.e.  $\bar{x}$  and  $S^2$  provide unbiased estimates of  $\xi$  and  $\sigma^2$ , respectively. It is also known that

$$\mathbf{D}\bar{x} = \frac{\sigma^2}{n} \quad (4)$$

and, if  $x$  is a Gaussian random variable,

$$\mathbf{D}S^2 = 2\sigma^4/(n-1). \quad (5)$$

Eq. (4) and (5) are used for evaluation of statistical errors.

In our MD simulations the random variable is the winding angle of one helical turn,  $\Phi$ , with expectation  $\mathbf{M}\Phi$  and variance  $\mathbf{D}\Phi$ . The torsional persistence length is computed as

$$l_t = L/\mathbf{D}\Phi$$

where  $L$  is the DNA length. It can also be obtained from the shifts in  $\mathbf{M}\Phi$  caused by external torques of different magnitude. The true values of  $\mathbf{M}\Phi$  and  $\mathbf{D}\Phi$  are estimated by using, respectively, the sample average and variance computed over all  $n$  points saved during an MD trajectory. However, Eq. (4) and (5) cannot be applied straightforwardly because they are valid only for statistically independent samples, i.e. the time intervals between the MD states must be suitably large compared

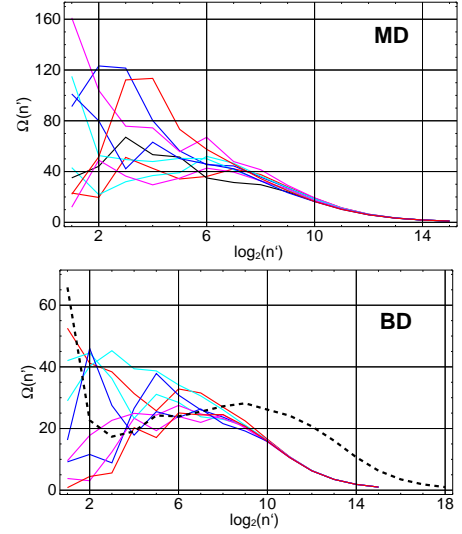


FIG. 3 FIG. 3S: Analysis of statistical errors by the method of block averages [41]. The upper panel displays the results for the nine MD trajectories of the GC-alternating fragment. For comparison, the lower panel presents a similar analysis of eight equivalent BD trajectories of the same DNA fragment. The equivalence means that MD and BD trajectories have identical durations and saving intervals. The BD simulations were carried out under zero torque by using a discrete WLC model from earlier reports [11, 36]. The black dashed line on the lower panel displays the results for a single eight times longer BD trajectory.

to the torsional relaxation time. The data saving interval is commonly much smaller, therefore, the errors are evaluated by using the method of block averages [41]. The trajectory is successively divided in  $n' = 2, 4, \dots, 2^{15}$  stretches (blocks) and the sample variances  $S'^2$  are computed by using  $n'$  block averages instead of individual samples. When the blocks are longer than the torsional relaxation time the block averages are independent and  $S'^2/n' \approx \text{const} = \sigma^2/\tilde{n}$ , where  $\tilde{n}$  is the effective number of independent samples provided by the trajectory. This value should be used in place of  $n$  in Eq. (4) and (5). In practice, it is convenient to draw the plots of

$$\Omega(n') = \frac{nS'^2}{n'\sigma^2},$$

with respect to  $\log_2 n'$ . When statistical independence is reached, such plots exhibit a plateau with  $\Omega \approx n/\tilde{n} = \tau^c$ , which gives the required estimate of  $\tilde{n}$ . Parameter  $\tau^c$  is the effective correlation time measured in trajectory saving steps.

Analysis of the MD data for the GC-alternating DNA is shown in Fig. 3S. For benchmark comparison, we also present a similar treatment of Brownian dynamics (BD) trajectories of a discrete wormlike chain (WLC) model used in our recent studies [11, 36]. With  $n'$  decreasing, the plots display emergence of a plateau and the growth of statistical noise. From these plots the  $\tau^c$  values are estimated as 50 and 30 for MD and BD, respectively, in



agreement with the corresponding relative rates of torsional relaxation [36]. The  $\tau^c \approx 50$  gives  $\tilde{n}=655$  and the relative error of 5.5% in the measured  $l_t$  values. This accuracy is sufficient for our purposes. The lower panel of Fig. 3S also displays the improvement that might be obtained with longer trajectories. As expected, with  $2^3$  longer BD trajectory the plateau is less noisy, but its value is similar.

#### Sequence-dependent distributions of twist fluctuations

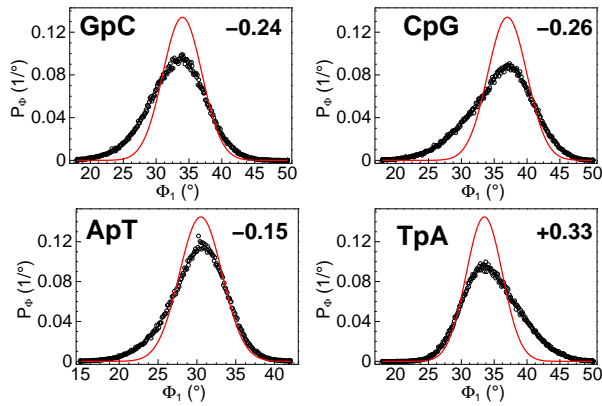


FIG. 4S: Normalized single-step probability densities of torsional fluctuations in MD. The red lines show the Gaussian distributions corresponding to the harmonic WLC model with the corresponding values of  $l_t$ . The Gaussians were shifted to the maxima of the computed distributions. The MD data shown are from the trajectories with zero applied torques.

Fig. 4S compares the observed single step distributions of torsional fluctuations in d(AT)<sub>7</sub> and d(GC)<sub>7</sub> with the corresponding distributions in equivalent coarse grained models, i.e. discrete WLC models with elastic parameters identical to those in MD. The numbers shown in the upper right corners display the corresponding mode skewness computed as

$$(\text{mean} - \text{mode})/(\text{standard deviation}).$$

Similar results for other base pair steps are shown in Fig. 5S. These data were obtained by MD simulations of tetradecamer fragments with the corresponding alternating or homopolymer sequences. Duration of all trajectories was about 8 ns.

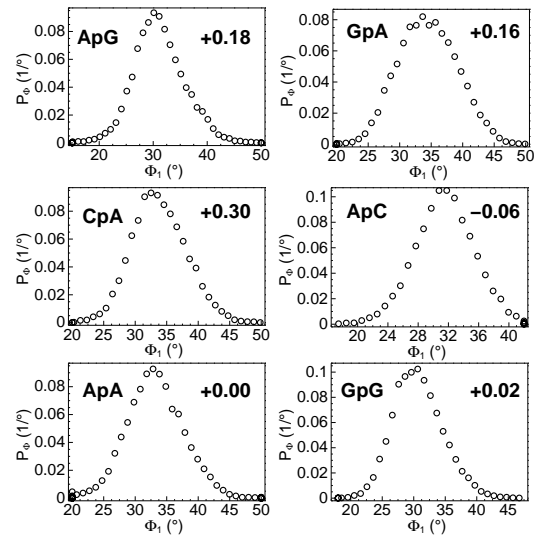


FIG. 5S: Normalized single-step probability densities of torsional fluctuations for other sequences.

\* Electronic address: alexey@ibpc.fr

- [1] A. V. Vologodskii and N. R. Cozzarelli, *Annu. Rev. Biophys. Biomol. Struct.* **23**, 609 (1994).
- [2] A. Travers and G. Muskhelishvili, *Nat. Rev. Microbiol.* **3**, 157 (2005).

- [3] B. J. Peter, J. Arsuaga, A. M. Breier, A. B. Khodursky, P. O. Brown, and N. R. Cozzarelli, *Genome. Biol.* **5**, R87 (2004).
- [4] J. A. Borowiec and J. D. Gralla, *J. Mol. Biol.* **195**, 89 (1987).
- [5] T. C. Boles, J. H. White, and N. R. Cozzarelli, *J. Mol. Biol.* **213**, 931 (1990).
- [6] L. Song, B. S. Fujimoto, P. G. Wu, J. C. Thomas, J. H. Shibata, and J. M. Schurr, *J. Mol. Biol.* **214**, 307 (1990).
- [7] P. R. Selvin, D. N. Cook, N. G. Pon, W. R. Bauer, M. P. Klein, and J. E. Hearst, *Science* **255**, 82 (1992).
- [8] A. N. Naimushin, J. B. Clendenning, U. S. Kim, L. Song, B. S. Fujimoto, D. W. Stewart, and J. M. Schurr, *Biophys. Chem.* **52**, 219 (1994).
- [9] B. J. Jordi, T. A. Owen-Hughes, C. S. Hulton, and C. F. Higgins, *EMBO J.* **14**, 5690 (1995).
- [10] K. L. Chow, M. E. Hogan, and R. J. Schwartz, *Proc. Natl. Acad. Sci. U. S. A.* **88**, 1301 (1991).
- [11] A. K. Mazur, *J. Chem. Theory Comput.* **5**, 2149 (2009).
- [12] S. Kannan, K. Kohlhoff, and M. Zacharias, *Biophys. J.* **91**, 2956 (2006).
- [13] J. Wereszczynski and I. Andricioaei, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16200 (2006).
- [14] G. L. Randall, L. Zechiedrich, and B. M. Pettitt, *Nucleic. Acids. Res.* **37**, 5568 (2009).
- [15] M. E. Hogan and R. H. Austin, *Nature* **329**, 263 (1987).
- [16] G. B. Koudelka, S. A. Mauro, and M. Ciubotaru, *Prog. Nucleic. Acid. Res. Mol. Biol.* **81**, 143 (2006).
- [17] P. J. Hagerman, *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265 (1988).
- [18] B. S. Fujimoto, G. P. Brewood, and J. M. Schurr, *Biophys. J.* **91**, 4166 (2006).
- [19] B. S. Fujimoto and J. M. Schurr, *Nature* **344**, 175 (1990).
- [20] T. R. Strick, D. Bensimon, and V. Croquette, *Genetica* **106**, 57 (1999).
- [21] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, *Nature* **424**, 338 (2003).
- [22] A. K. Mazur, *Biophys. J.* **91**, 4507 (2006).
- [23] A. K. Mazur, *Phys. Rev. Lett.* **98**, 218102 (2007).
- [24] S. Arnott and D. W. L. Hukins, *Biochem. Biophys. Res.*

- Commun. **47**, 1504 (1972).
- [25] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
  - [26] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
  - [27] T. E. Cheatham, III, P. Cieplak, and P. A. Kollman, J. Biomol. Struct. Dyn. **16**, 845 (1999).
  - [28] W. L. Jorgensen, J. Chandreskhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
  - [29] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
  - [30] A. K. Mazur, J. Comput. Chem. **18**, 1354 (1997).
  - [31] A. K. Mazur, J. Chem. Phys. **111**, 1407 (1999).
  - [32] A. K. Mazur, J. Am. Chem. Soc. **120**, 10928 (1998).
  - [33] A. K. Mazur, J. Phys. Chem. B **102**, 473 (1998).
  - [34] X.-J. Lu and W. K. Olson, Nucleic Acids Res. **31**, 5108 (2003).
  - [35] A. K. Mazur, J. Phys. Chem. B **112**, 4975 (2008).
  - [36] A. K. Mazur, J. Phys. Chem. B **113**, 2077 (2009).
  - [37] A. K. Mazur, J. Chem. Theory Comput. **1**, 325 (2005).
  - [38] S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, III, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, et al., Biophys. J. **89**, 3721 (2005).
  - [39] A. Perez, I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, and M. Orozco, Biophys. J. **92**, 3817 (2007).
  - [40] X.-J. Lu and W. K. Olson, J. Mol. Biol. **285**, 1563 (1999).
  - [41] D. Frenkel and B. Smit, *Understanding Molecular Simulations. From Algorithms to Applications* (Academic Press, New York, 1996).